**SENTIMENT ANALYSIS OF CUSTOMER FEEDBACK ON RESTAURANTS USING IBM CLOUD**

Mini Project documentation submitted to

## JAWAHARLAL NEHRU TECNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfillment of the requirements for the award of the degree of

## BACHELOR OF TECHNOLOGY

In

## COMPUTER SCIENCE AND ENGINEERING

Submitted By

## 1. P.SADVISHA 17UK1A0583

## 2. AV.AKSHAY 17UK1A05A7

**3. O.LAXMAN 17UK1A0522**

**4.MD.MOINUDDIN 19UK5A0517**

**5. V.SEETHARAM 19UK5A0513**

Under the guidance of

#### Dr.Rakesh Nayak

Assistant professor

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

## VAAGDEVI ENGINEERING COLLEGE

Affiliated to JNTUH, HYDERABAD

BOLLIKUNTA, WARANGAL (T.S) – 506005 2018-2022

Affiliated to JNTUH, HYDERABAD BOLLIKUNTA, WARANGAL (T.S) – 506005 2018-2022

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING VAAGDEVI ENGINEERING COLLEGE

## VCE-Vaagdevi College of Engineering WARANGAL

**CERTIFICATE**

This is to certify that the Mini Project report entitled **“ SENTIMENT ANALYSIS OF CUSTOMER FEEDBACK ON RESTAURANTS USING IBM CLOUD*”*** is being submitted by Sadvisha(17UK1A0583),AV.Akshay(17UK1A05A7),Laxman(17UK1A0522),MD.Moinuddin   
(19UK1A0517),V.Seetharam(19UK1A0513) in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2018- 2022.

#### 

#### Dr.Rakesh Nayak Dr. R. NAVEEN KUMAR

Guidence Head of the department

# EXTERNAL

## ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr. P. Prasad Rao**, Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this Mini Project in the institute.

We extend our heartfelt thanks to **Dr. R. Naveen Kumar**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the Mini Project.

We express heartfelt thanks to the Mini Project coordinator,Ms. G. Arunakranthi, Assistant Professor, Department of CSE for her constant support and giving necessary guidance for completion of this Mini Project.

We express heartfelt thanks to the the guide **Dr. Rakesh Nayak**, Assistant Professor, Department of CSE for his constant support and giving necessary guidance for completion of this Mini Project.

We express our sincere thanks and gratitude to **The Smart Bridge**, for providing an internship. Finally, we express our sincere thanks and gratitude to our family members, friends for their encouragement and outpouring their knowledge and experiencing

**ABSTRACT**

Today, digital reviews play a pivotal role in enhancing global communications among consumers and influencing consumer buying patterns. E-commerce giants like Zomato, Swiggy etc provide a platform to consumers to share their experience and provide real insights about the performance of the product to future buyers. In order to extract valuable insights from a large set of reviews, classification of reviews into positive and negative sentiment is required. Sentiment Analysis is a computational study to extract subjective information from the text. In the proposed work, over 4,000,00 reviews have been classified into positive and negative sentiments using Sentiment AnalysisThe growth of web contributes a huge quantityof user created content such as customer feedback,opinions and reviews. Sentiment analysis in web embraces.The problem of aggregating data in the web and extractionabout opinions. Studying the opinions of customers helpsto determine the people feeling about a product and how itis received in the market. Various commercial tools areavailable for sentiment analysis. In this paper, we proposea system which classifies the reviews on a scale of good or bad based on the sentiments in the words. The groups of wordsused to make a decision to rate the reviews are displayedas word cloud.

# TABLE OF CONTENTS

1. **INTRODUCTION…………………………………………………………………………….6**
   1. **OVERVIEW………………………………………………………………………….......6**
   2. **PURPOSE…………………………………………………………………………….......6**
2. **LITERATURESURVEY………………………………………………………….………....7**
3. **THEORITICAL ANALYSIS………………………………………………………….......8**
   1. **BLOCK DIAGRAM…………………………………………………………………….8**
4. **EXPERIMENTAL INVESTIGATION………………………………………………….….9**
   1. **METHODOLOGY……………………………………………………………………..…5**
      1. **DATA COLLECTION……………….…………………………………..…...9**
      2. **DATA PREPROCESSING ………………………………………………….10**
      3. **SENTIMENTAL ANALYSIS…………..……………………………………10**
      4. **CLASSIFICATION…………………………………………………………...11**
5. **FLOW CHART………………………………………………….…………………………….12**
6. **RESULTS & DISCUSSION………………………………………………………………….13**
7. **ADVANTAGES…………………………………………………………………………….…14**
8. **DISADVANTAGES……………………………………………………………………..........14**
9. **APPLICATIONS………………………………………………………………………..........15**
10. **CONCLUSION……………………………………………………………….…………........15**

**11 BIBILOGRAPHY…………………………………………………………………………....16**

1.INTRODUCTION

**1.1 Overview**

Sentiment analysis is a huge volume increasing at a humongous rate everyday which has made it almost impossible to evaluate the data manually. In Social media, twitter, restaurant site people share their opinion as in a huge number of their prevalence. In order to make the process of analyzing the text automatic there are various machine learning techniques that could be applied. The data set is for those enthusiasts who are willing to play with text data and perform sentiment analysis or text classification. The huge quantity of data in textual is generated every day has no value unless processed. The text data issue can be resolved by adopting Data mining technique using r tool. Our experimental work intends to adopt data mining techniques for the effective prediction of text data. This data set consists of actual reviews from real people. So this data set will give a real time experience as to how todeal with textual data.

**1.2 Purpose**

Most of the customers will follow and choose the best restaurants on the basis of reviews and ratings. So reviews Play a Crucial Role in any model or system. The approach to this problem is based on review text content analysis and uses the principles of natural language process (the NLP method) and Machine learning. After applying the above method we can classify whether it is a positive review or negative review and can also visualize the total no of positive reviews and negative reviews. We are working on developing an algorithm that can help in classifying the reviews on the basis of positive and negative reviews with the help of a predefined dictionary of words.

A web Application is created where user can enter their feedback, the entered text is analyzed by the model built and prediction is showcased on UI.

**2.LITERATURE SURVEY**

“What others think? It is always important information in adecision-making process. Every day people discuss variousproducts on social media sites. Web and its associateddistribution services provides information services such asonline services where data objects are linked together tofacilitate interactive access. Web pages may not have apredefined schema or pattern and it is difficult for computersto understand the semantic meaning. Companies want a pieceof that pie to determine how their audience communicates tofind the important information that drives business. Sentimentanalysis is the robotic mining of opinions and feelings fromcontent through Natural Language Processing (NLP).Sentiment analysis is nothing but categorizing opinions in thegiven content or documents into "positive" or "negative or"neutral"Sentiment analysis can be processed in three levels: aspect-

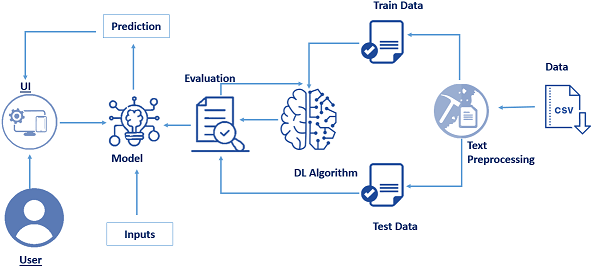
level, document-level, and sentence-level. Sentiment analysis in document level considers the entire document as a singletopic and classifies positive or negative sentiment. Thesentiment expressed in each and every sentence is classified insentence level. Since sentences are part of the documents doesnot make much difference between sentence and documentlevel. To get the detail opinion or sentiments have to processthe document to aspect level. Aspect level sentiment analysisis to determine the features of the sentiment conveyed towardseach aspect and the given target entities. This proposed modelhandles sentiment polarity classification which is afundamental problem of sentiment classification. The onlinedata have several drawbacks to hinder the sentiment analysis

task. The first defect is anybody can post their own contentsand quality or impact of their comment is not assured. Onlinespammers may post their fake opinions The next fault isthe polarity of reviews cannot be ascertained or unavailable.The dataset used in this paper is around 3000 food reviews collected from Zomato. Each post in Zomato is inspected andverified by the company before it gets posted. Each review hasa rating scale from 1 to 5 stars which can be used to identifythe sentiment polarity.This is quite widely used, bothin the research NLP community and also among commercial and government users of open source NLP technology. The method suggest that it follows from a simple, approachable and easy design, which can be

Straightforward interfaces, the inclusion of robust and in the good quality analysis components, and not to requiring use of a large amount of associated baggage.

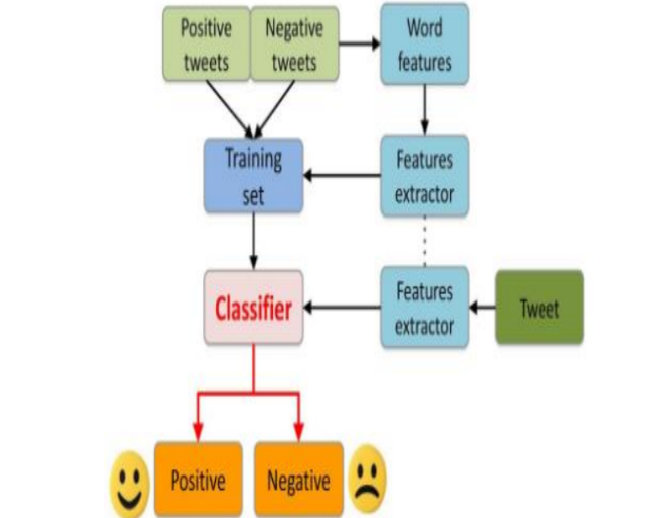
**3.THEORETICAL ANALYSIS**

**3.1 Block Diagram**

****

**4.METHODOLOGY**

This proposed work is to predict the text automatically based on the data set values stored by using the r tool. Byusing the training data set values it is possible to predict the text data.



 The Fig 1 depicts the architecture of the proposed model used in the prediction of Sentiment analysis. It consists of 4 steps :

**A. Data Collection**

In this step data is taken out from kaggle in a recognized format. Missing fields are evacuated in this process & thus

the data is transformed. Sentiment Analysis can be considered a classification process. There are three main classification levels in sentiment analysis document-level, sentence-level, and aspect-level sentiment analysis. Level of document it aims to classify an opinion document which as a positive or negative opinion expression. It considers the full document as a basic information unit.

**B. Data Prep-processing**

The collected raw data of restaurant reviews consist of large number of attributes and also there will be missing values. The reducing the attributes is required,extracting the required attributes is also much essential. So inorder to get importance of the each variable or attributes migrittr algorithm is applied. Migrittr algoirithm which selects the attributes based on predictor, here predictor consisdered restaurant review. Feature or Attribute extraction is done using migrittr algorithm. In detail steps working of migrittr algorithm.

In Data cleaning once attributes are removed,filling the missing values, removing inconsistent datameasuring the central tendency for the attribute such as mean median, quartile is done. In data preprocess the data is cleaned andthe extracted data before analysis. Non-textual contents and contents that are irrelevant for the analysis are identified and eliminated.

**C. Sentiment Analysis**

The reviews sources are mainly review sites. Sentiment analysis is not only applied on product reviews but can also be applied on stock market, news articles, or political debates. In political debates for example, we could figure out people’s opinions on a certain election candidates or political parties. The election results can also be predicted from political posts. The sites like social media and micro

blogging sites are taken a very good source of information because many people share and discuss their opinions about positive and negative opinion freely.

**D. Classsification**

The lexicon-based approach is to finding the opinionmining which is used to analyze or to predict the text.There are two methods in this approach. The dictionary-based approach which depends on finding opinion seedwords, and then searches the dictionary of their synonymsand antonyms. The corpus-based approach begins with aseed list of opinion words, and then finds other opinionwords in a large corpus to help in finding opinion wordswith context specific orientations. This could be done by

using statistical or semantic methods.Data mining has got two most frequent modelling goals –classification & prediction. Classification model classifies

discrete, unordered values or data. In this predictionprocess, the classification techniques utilized are, naivebayes classifier.

Fig1: Architecture of the proposed modelA. Naïve Bayes.It is one of the popular classification techniques ofalgorithms used in data mining. It is a probability classifier.It links the attributes mutually & is dependent on thenumber of parameters. The principle here is that thevariables provided are independent. It generates accurateresults with appropriate calculation & provides fast results.

It is based on Bayes theorem & the formula is, P (label|features) = P (label)\* P(features|label) P(features) ….(eq1)

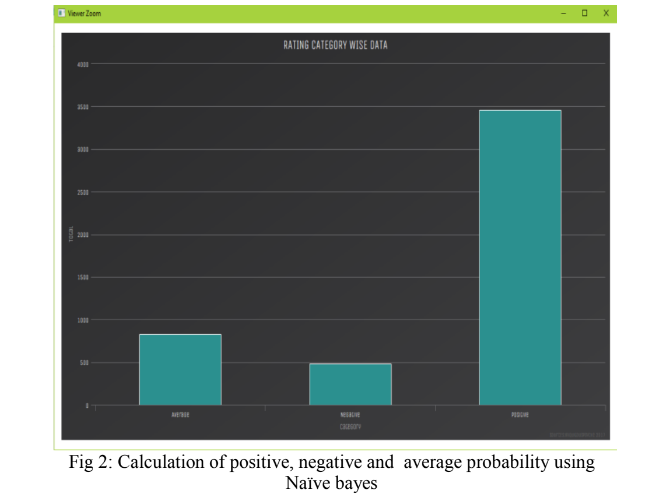
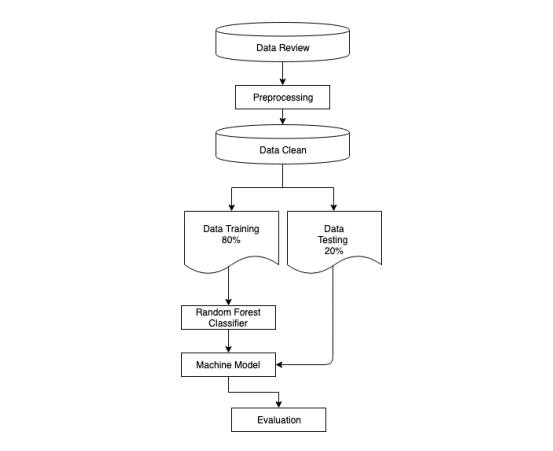


Fig 2 shows the graphical representation of the result.The graph is generated by using the hchart chart.X-axis is the number data set values and Y- axis represents

the Probability of positive or negative with the percentagevalues.

Fig 2: Calculation of positive, negative and average probability usingNaïve bayes.

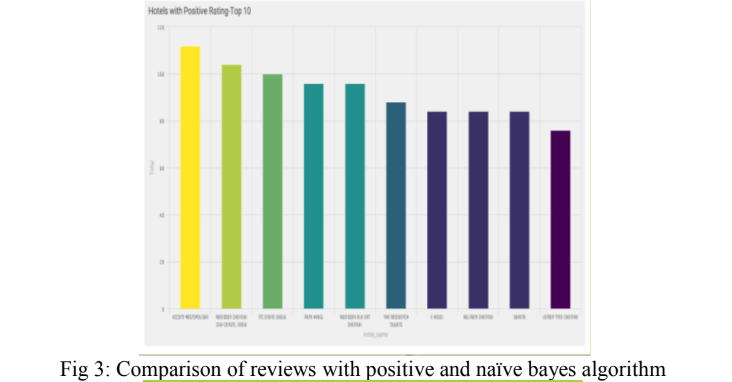
**4. Flow Chart**

****

**5. RESULT & DISCUSSION**

Data analysis is the most crucial part of any proposedwork. Data analysis summarizes collected data. It involvesthe interpretation of data gathered through the use ofanalytical and logical reasoning to determine patterns,relationships or trends. To examine critically and to bringout the essential elements or give the essence to analyze data. To examine carefully and in detail so as to keyfactors, possible results. Following snapshots shows the

results obtained in each step of the process.



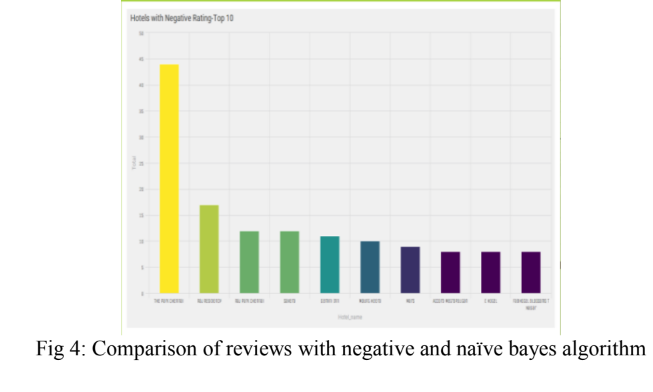


Fig 3 depicts naïve bayes algorithms with the accuracyvales like positive is 3456, negative is 485, and average is27 . Fig 4 shows the test cases of the algorithms withdifferent accuracy values like 112. and Fig 5 shows onemore test case result like 44 and 28.

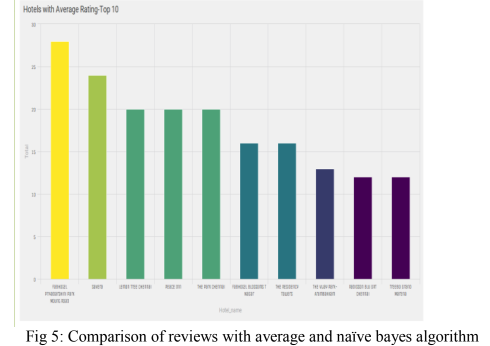


Fig 3: Comparison of reviews with positive and naïve bayes algorithm

Fig 4: Comparison of reviews with negative and naïve bayes algorithm

**6. ADVANTAGES :**

By using sentiment analysis, you gauge how customers feel about different areas of your business without having to read thousands of customer comments at once.

If you have thousands of feedback per month, it is impossible for one person to read all of these responses. By using sentiment analysis and automating this process, you can easily drill down into different customer segments of your business and get a better understanding of sentiment in these segments.

**DISADVANTAGES:**

Flaws of using sentiment analysisWhile sentiment analysis is useful, it is not a complete replacement for reading survey responses. Often, there are useful nuances in the comments themselves. Where sentiment analysis can help you further is by identifying which of these comments you should read.

**7. APPLICATIONS**

1.Intent Analysis

2.Context Semantic Search

3.Social media monitoring

4.Customer support

5.Customer feedback

6.Brand monitoring and reputation management

7.Voice of customer (VoC)

8.Voice of employee

9.Product analysis

10.Market research and competitive research

**8. CONCLUSION**

The proposed work starting from the analysis of differentstudies provided in the literature, provides a classificationof sentiment classification approaches with respect tofeatures/techniques and advantages /limitations, tools forsentiment analysis with respect to the different techniquesused for sentiment analysis.The sentiment classification approaches can beclassified in machine learning, lexicon based and hybridapproach. The machine learning approach is used forpredicting the sentiments based on trained and test datasets. In our lexicon based approach does not need anyprior training in order to mine the data.

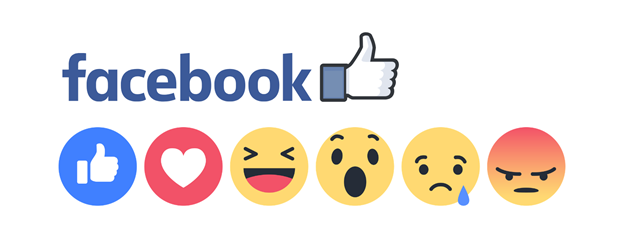
**9. FUTURE SCOPE**

In the future, **sentiment analysis** will delve deeper, beyond the concept of the number of reactions, likes, comments, and shares in a post, to reach and comprehend the significance of social media conversations and what they reveal about consumers.

As a result, sentiment analysis tools are becoming necessary for these businesses to survive in such a competitive market.

Even today, the majority of sentiment analysis in any project is carried out almost entirely by corporations and brands using data from social media, survey responses, and other sources of user-generated content.

## And, by examining and evaluating customer sentiments with such tools, these brands are able to gain an in-depth understanding of consumerbehaviours and, as a result, better serve their audiences with the products, services, and experiences they provide.



For example,Facebook had replaced its only button, called reactions, with a slew of emojis. This enabled the user to express their emotions more freely.As a result, every time the major social media platforms update and add new features, the data underlying those interactions becomes broader and deeper.

**9. BIBILOGRAPHY**

1]Mikel Joaristi, Edoardo Serra, Francesca Spezzano “Evaluatingthe Impact of Social Media in Detecting the Restaurants Violatingthe Health Norms” In 2016 IEEE/ACM International Conferenceon Advances in Social Networks Analysis and Mining(ASONAM).

2] J. P. Schomberg, O. L. Haimson, G. R. Hayes, and H. Anton-Culver. Supplementing public health inspection via social media.PLoS ONE, 11(3), 03 2016.

3] A. Sadilek, S. Brennan, H. Kautz, and V. Silenzio. nemesis:Which restaurants should you avoid today? In HCOMP, 2013.

4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion,O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, etal. Scikit-learn: Machine learning in python. The Journal ofMachine Learning Research, 12:2825–2830, 2011.

5] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard,and D. McClosky. The stanford corenlp natural languageprocessing toolkit. In ACL (System Demonstrations), pages 55–60, 2014.

6] K. Lee, A. Agrawal, and A. N. Choudhary. Mining social mediastreams to improve public health allergy surveillance. InASONAM, pages 815– 822, 2015.

7] K.Lee,A.Agrawal, andA.Choudhary. Real-time diseasesurveillance using twitter data: demonstration on flu and cancer.In KDD, pages 1474–1477, 2013.

**APPENDIX**

1. Source Code.